International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12P 19/34, C07H 21/00 | A1

(11) International Publication Number: WO 92/07948

(43) International Publication Date: 14 May 1992 (14.05.92)

(21) International Application Number: PCT/US91/08233

(22) International Filing Date: 5 November 1991 (05.11.91)

(30) Priority data:

610,973	6 November 1990 (06.11.90)	US
737,919	29 July 1991 (29.07.91)	US

(71) Applicant: THE LUBRIZOL CORPORATION [US/US]; 29400 Lakeland Boulevard, Wickliffe, OH 44092-2298 (US).

(72) Inventors: CARDINEAU, Guy, A. ; 1013 Tramore Trail, Madison, WI 53717 (US). FILNER, Philip ; 7717 Glen Road, Danville, CA 94526 (US).

(74) Agents: WESEMAN, James, C. et al.; Limbach & Limbach, 2001 Ferry Building, San Francisco, CA 94111 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published

With international search report.

Title: COMPOSITIONS AND METHODS FOR ANALYZING GENOMIC VARIATION

(57) Abstract

Compositions and methods are described for analyzing genomic variation involving single primer amplification and detection of polymorphisms without the need for digesting nucleic acid with restriction enzymes or transferring nucleic acid for hybridization.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE+	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

Description

Compositions and Methods for Analyzing Genomic Variation

Field of the Invention

The present invention relates to compositions and methods for detecting and analyzing genomic variation, and, more particularly, for detecting and analyzing nucleic acid polymorphisms.

Background of the Invention

Application of modern nucleic acid manipulation techniques allows for the detection of discrete nucleic acid sequences within a complex genome. In particular, utilization of immobilization techniques, such as those described by Southern, J. Mol. Biol. 98:503 (1975), in combination with restriction enzymes, has allowed for the identification by hybridization of genes or gene fragments among a mass of fractionated, genomic DNA.

Restriction enzymes are endonucleases which recognize a specific base sequence (site) of a double-stranded nucleic acid molecule and catalyzing cleavage of the molecule at a precise location. See e.g., Smith and Wilcox, J. Mol. Biol. 51:379 (1970). Restriction sites provide a convenient means for fragmenting DNA into pieces of different length. Once the DNA has been cleaved with a restriction enzyme, the various fragments can be separated by size using gel electrophoresis and transferred to DNA-binding membranes such as nitrocellulose by blotting. Thereafter, a short single-stranded DNA sequence that is complementary to a target sequence in the fragments, i.e. a probe, can be labelled and used to detect the target sequence by hybridization.

It is well known that there may be one or more alternate forms of a gene occupying a given locus on

different versions of the same chromosome. Such genomic variations, both between individuals within a species and between species, occurs by processes such as recombination and mutation. There may be one or more base changes (substitutions, additions or deletions), repeat sequences, and even changes in the gene copy number. There also may be variation in the sequences to the left or right of a gene, i.e. variation in "flanking" sequences.

The use of restriction site differences has been a favored means for characterizing genomic variation. The procedure uses the techniques developed by Southern, described above, except that genomic DNA from distinct sources is compared (e.g. different individuals within a species, individuals from different species, etc.). The variability of restriction sites in gene and gene-flanking sequences results in DNA fragments of different length (and consequently different migration rate on gels). The hybridization of probes to specific restriction fragments serves to identify genetic loci or other sequences of interest. Using a panel of different restriction enzymes, it is possible to distinguish the DNA samples from the various distinct sources on the basis of their hybridization banding pattern on Southern blots. See e.g. G. Vassart et al., Science 235:683 (1987). M. Georges et al., JAVMA 193:1095 (1988). M. Georges et al., Cytogenet. Cell. Genet. 47:127 (1988). Where a probe hybridizes to restriction fragments of different lengths to create different hybridization banding patterns for different samples, the variation is referred to as Restriction Fragment Length Polymorphism (RFLP).

RFLP analysis provides a means for following the segregation of genes derived from each parent. Indeed, the segregation and intensity of

hybridization can be used to estimate the copy number, linkage organization, and complexity of closely related genes that are otherwise difficult to assess. RFLP loci can be mapped genetically by determining the frequencies of recombination between different RFLP loci. By correlating the segregation of the RFLP loci with the segregation of known genetic markers, the genetic maps of RFLP markers can be merged with genetic maps of other types of markers, e.g., morphological or biochemical markers.

There are some distinct disadvantages to the current RFLP technique. First, it is very labor intensive; many steps are required to take genomic DNA to the point where RFLPs can be recognized. Second, it is slow; because of the gel transfer and hybridization steps, results can take a number of days. Finally, it is expensive; the cost for labor and restriction enzymes required for RFLP analysis can be prohibitive for large breeding studies.

In view of the disadvantages to current RFLP techniques, other techniques for studying genomic variation have been sought. One approach to generating multiple nucleic acid bands involves the use of nucleic acid amplification. For example, J.S. Chamberlain et al., Nucleic Acids Research 16:11141 (1988) have described an amplification method to amplify more than a single nucleic acid sequence. The approach involves using more than one pair of primers in a Polymerase Chain Reaction. The Polymerase Chain Reaction ("PCR) is described by K.B. Mullis et al., U.S. Patent Nos. 4,683,195 and 4,683,202, hereby incorporated by reference. PCR is a procedure for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification by introducing a large excess of two oligonucleotide primers to the

DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are selected complementary to flanking sequences on respective strands of the double-stranded target sequence. To effect amplification, the mixture is denatured and the primers are then allowed to anneal to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (i.e. denaturation, annealing and extension constitute one "cycle;" there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the two primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process and the use of the amplified sequences as templates for further amplification in subsequent cycles, the method is referred to by the inventors as the "Polymerase Chain Reaction" (hereinafter PCR). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR-amplified".

As noted above, the PCR procedure uses two primers with the intent of amplifying a single sequence whose length is defined by the position of the primers. J.S. Chamberlain et al., supra, used more than one pair of primers in order to produce multiple sequences. The use of multiple pairs of primers (so-called "multiplex genomic DNA

amplification") for analysis of genomic variation also has distinct disadvantages. First, the technique requires knowledge of the target sequence of interest, i.e. the sequence in the genome to be amplified. Secondly, the addition of each extra primer pair frequently requires modification of primer annealing temperatures, time of annealing, polymerase extension times, and the amount of enzyme required. Modifying the primer annealing temperatures may change the stringency of hybridization which, in turn, may result in amplification of sequences having little homology to the primers and this may severely complicate an analysis of genomic variation (amplification of sequences of different length in two individuals would not necessarily indicate recombinant or mutational variation).

A.J. Jeffreys et al., Nucleic Acids Research 16:10953 (1988) describe a procedure where PCR is used to produce multiple sequences. The procedure uses two primers directed at the known flanking sequences of hypervariable, tandem-repeated ("minisatellite") loci with the intent of amplifying the entire minisatellite. The procedure has a number of disadvantages. First, the procedure requires that the hypervariable "minisatellite" regions first be identified. See A.J. Jeffreys et al., Nature 314:67 (1985). Second, the procedure requires specific polynucleotide probes. See A.J. Jeffreys patents: U.K. Patent 2166445 and EPC 0238 329. Thirdly, minisatellite PCR must be terminated before the yield of product reaches chemical amounts because of production of a heterodisperse smear. Thus, a signal generation step like hybridization must be used to identify the products.

D. L. Nelson et al., Proc. Nat. Acad. Sci. USA 86:6686 (1989) describe a procedure where PCR is used to produce multiple sequences. See also S.A. Ledbetter et al., Genomics 6:475 (1990). The procedure uses single primers directed at the known sequences of short interspersed repeats believed to exist in great number (approximately 900,000 in the haploid human genome). This procedure also has a number of disadvantages. First, this procedure, like the minisatellite PCR procedure, requires that the primers and the sequences to be amplified be known. Secondly, while PCR can be run to yield chemical amounts of discrete product, this is only possible if less than the entire genome is used. When the entire genome is used, there is again the production of a smear.

The present invention involves a more desirable means of amplifying more than a single sequence. The present invention provides a method of obtaining information as important and useful as RFLP data, but without the accompanying labor, time and expense. The present invention does not require the use of restriction enzymes or nucleic acid transfer to perform an analysis of genomic variation. Furthermore, while the invention involves amplification, there is no need to modify annealing temperatures. Finally, the invention allows for amplification in chemical amounts utilizing nucleic acid representing the entire genome.

Objects and advantages other than those above set forth will be apparent from the following description when read in connection with the accompanying figures.

Summary of the Invention

The present invention relates to compositions and methods for detecting and analyzing genomic variation and, in particular, nucleic acid polymorphisms. In one aspect, the invention comprises a method for amplifying a plurality of sequences found in a nucleic acid sample comprising providing a nucleic acid sample comprising nucleic acid sequences of a distinct nucleic acid source, providing a primer comprising an oligonucleotide sequence of at least eleven nucleotides which is capable of hybridizing to at least a portion of the nucleic acid sequences and generating a plurality of amplification products therefrom in an amplification system, thereby amplifying at least one portion of the nucleic acid sequences, so that discrete portions of the nucleic acid sequences are detectable.

In accordance with another aspect of the subject invention, a method for amplifying a plurality of sequences found in a nucleic acid sample is provided which comprises providing a nucleic acid sample comprising single-stranded nucleic acid sequences from a distinct nucleic acid source, bringing the sample together with a primer comprising an oligonucleotide sequence of at least eleven nucleotides capable of hybridizing to a plurality of regions within the nucleic acid sequences, under conditions which synthesize an extension product which is complementary to a portion of the nucleic acid sequences, and wherein the extension product, when separated from the nucleic acid sequence, is also capable of hybridizing to the oligonucleotide primer, separating the primer extension products from the nucleic acid sequences to which they were hybridized to form single-stranded templates, and bringing together the templates with the oligonucleotide primer under conditions which

synthesize a primer extension product from the primers hybridized to the templates, thereby amplifying a plurality of portions of the nucleic acid sequences, so that discrete portions of the nucleic acid sequences are detectable.

In accordance with another aspect of the subject invention, a method for detecting variation between nucleic acid samples is provided comprising providing at least two nucleic acid samples each comprising nucleic acid sequences representing the entire genome of a distinct nucleic acid source; providing a single primer comprising an oligonucleotide sequence of at least eleven nucleotides capable of hybridizing to at least a portion of the nucleic acid sequences from each source and capable of generating a plurality of amplification products therefrom in an amplification system; and bringing together each nucleic acid sample with the primer in a separate amplification system, thereby amplifying at least one portion of the nucleic acid sequences from each source, so that discrete portions of the nucleic acid sequences are detectable. Certain embodiments of this aspect of the invention also provide that the amplified portions of nucleic acid sequences from each nucleic acid source so amplified are compared to determine the degree of homology between the nucleic acid sources.

Additional aspects of the invention provide primers which find use in the present method, and a mixture of nucleic acid sequences comprising a plurality of double-stranded products comprising a first and a second single-stranded polynucleotides each having a 5'-terminal region sequence and a 3'-terminal region sequence which is substantially the inverted complement thereof, and an internal region.

Brief Description of the Figures

Figure 1A is a photograph of an ethidium bromide stained gel of electrophoresed SPAR-amplified and PCR-amplified sequences from chicken genomic DNA. Figure 1B is a photograph of an autoradiograph of the electrophoresed SPAR-amplified and PCR-amplified sequences of Figure 1 following transfer to nitrocellulose and hybridization with a radioactive probe.

Figure 2 is a photograph of an ethidium bromide stained gel of the electrophoresed SPAR-amplified sequences from chicken genomic DNA amplified primers of the present invention.

Figure 3 is a photograph of an ethidium bromide stained gel of the electrophoresed SPAR-amplified sequences from chicken genomic DNA under three annealing temperatures (Figure 3A, 3B and 3C).

Figure 4 is a photograph of an ethidium bromide stained gel of the electrophoresed, SPAR-amplified sequences from chicken genomic DNA derived from different sources, i.e. different individual chickens.

Figures 5A and 5B are (direct print) photographs of ethidium bromide stained gels of the electrophoresed, SPAR-amplified sequences from corn genomic DNA derived from different sources, i.e. different individual corn plants, each from a different inbred corn line. (Direct printing results in light bands on a dark background).

Detailed Description of the Invention

The present invention relates to compositions and methods for detecting and analyzing genomic variation and, in particular, nucleic acid polymorphisms.

In accordance with the present invention, a method is provided for amplifying a plurality of nucleic acid sequences found in a nucleic acid sample. In the practice of this method, a nucleic acid sample is employed which comprises nucleic acid sequences of a distinct nucleic acid source. Any source of nucleic acid, in purified or non-purified form, can be utilized as the source for the nucleic acid sample, provided the sample is from a distinct nucleic acid source which is suspected of harboring polymorphisms useful for detecting genomic variation.

Thus, the method may employ any nucleic acid, for example DNA or RNA, including messenger RNA, which may be single-stranded or double-stranded. In addition, a DNA-RNA hybrid which contains one strand of each distinct nucleic acid may be utilized. It is also possible to utilize a mixture of any one or more of such nucleic acids provided they are from a source appropriate to facilitate the analysis of genomic variation.

In the practice of the invention, the nucleic acid sample does not need to be provided in pure form; it may be a fraction of a more complex mixture, e.g., it may constitute only a minor fraction of a particular sample of biological origin.

In certain embodiments, the present invention contemplates that the nucleic acid is derived from a microorganismal source, such as virus, bacteria, fungi, yeast, algae, mycoplasma and protozoa.

In other embodiments, the present invention contemplates that the nucleic acid provided is derived from a plant. Representative examples of plants which serve as nucleic acid sources include the angiosperms as well as the gymnosperms. Examples of angiosperms include both monocotyledons and dicotyledons, such as corn, barley, wheat, apple,

alfalfa, soybean, oil rape, tobacco and tomato. Examples of gymnosperms include cycads and conifers, such as loblolly pine.

In additional embodiments, the present invention employs nucleic acid samples derived from animal sources, including both vertebrates and invertebrates. Representative examples of vertebrates include mammals, birds, reptiles and amphibians, such as human, horse, dog, cow, chicken, mouse, rat and salmon. Representative examples of invertebrates include arthropods, mollusks, flatworms, annelids and echinoderms.

In practicing the present method, a primer is provided capable of hybridizing to at least a portion of the nucleic acid sequences in the nucleic acid sample and generating a plurality of amplification products from the nucleic acid sequences in an appropriate amplification system.

As used herein, the term primer refers to an oligonuclectide, whether occurring naturally (e.g., as a component of a purified restriction enzyme digest product) or constructed synthetically. The term oligonucleotide is defined as a molecule comprised of two or more nucleotides e.g., deoxyribonucleotides or ribonucleotides.

The primers which find use in the present invention will form extension products in an amplification system. It is not intended that the present invention be limited by the mechanism(s) whereby the primers of the present invention form extension products. For example, the primer may be substantially complementary to at least a portion of the nucleic acid sequences contained in the nucleic acid sample, so that the primer is capable of hybridizing to the sequences and forming temporarily stable complexes in the reaction conditions utilized

in the selected amplification system. As used herein, the portion of the nucleic acid sequence to which the primer is capable of hybridizing and forming a temporarily stable complex is termed the template. It may be that the primer has sufficient complementarity with a template in the sample to hybridize therewith and form extension products in an amplification system. As used herein, an extension product is the collection of generated nucleic acid sequences which contain a particular primer sequence together with an additional sequence complementary to the nucleic acid sequence bordering the template.

In accordance with the invention, a primer will comprise an oligonucleotide capable of hybridizing to at least a portion of the nucleic acid sequences and capable of generating a plurality of amplification products therefrom in an amplification system. In presently preferred embodiments, the selected primer will be capable of hybridizing to a plurality of regions within nucleic acid sequences in the nucleic acid sample, and will be capable of hybridizing to an extension product which contains the oligonucleotide sequence and is substantially complementary to a portion of the nucleic acid sequences.

Particular primers will find use in the present invention, including oligodeoxyribo-nucleotides generally in accordance with the formula:

$$5'\text{-}X_1\text{-}X_2\text{-}X_3\text{-}X_4\text{-}G\text{-}A\text{-}C\text{-}Y_1\text{-}Y_2\text{-}Y_3\text{-}3' \qquad (I)$$

wherein

X_1 is desX_1 or an oligonucleotide of from 3 to 11 bases selected from the group consisting of A-G-A-G, G-A-C-C-A-A-C-T-G-G-T, and C-C-C;

X_2 is desX_2 or an oligonucleotide of 3 bases selected from the group consisting of A-A-T and C-C-C;

X_3 is desX_3 or an oligonucleotide of from 1 to 3 bases selected from the group consisting of T-C-T, G-G-T, A-A-C, and G;

X_4 is an oligonucleotide of 3 bases selected from the group consisting of C-A-A, A-G-C and A-C-A;

Y_1 is a base selected from the group consisting of C and A;

Y_2 is an oligonucleotide of 3 bases selected from the group consisting of G-G-C, C-C-T, G-A-C, and T-A-C; and

Y_3 is desX_3 or an oligonucleotide of from 5 to 13 bases selected from the group consisting of A-A-C-A-G-G, G-G-G-C-C-T-G-G-T-C-G-A-T, A-G-A-C-A and T-T-C-C-C-C-C;

with the proviso that at least approximately fifty percent (50%) of the deoxyribonucleotides of the primer are deoxyguanylic acid and deoxycytidylic acid.

As used herein, the term "des-" is taken to mean that the particular base or oligonucleotide fragment may not be present in selected primer sequences within the scope of the formulae.

More usually, primers having sequences in accordance with formula I will have the formula:

$$5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-Y_2-Y_3-3' \quad (II)$$

wherein X_1, X_2, X_3, Y_2 and Y_3 are as previously defined.

Presently preferred primers in accordance with formulas I and II include primers of the formula:

$$5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-G-(G/A)-C-Y_3-3' \quad (III)$$

wherein X_1, X_2, X_3 and Y_3 are as previously defined. Of particular interest among the presently preferred primers of formula III are those selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'

5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'

5'-GACCAACTGGTAATGGTAGCGACCGGC-3'

5'-GACAGACAGACAGACA-3' and

5'-CCCCCCAACCAAGACCTACTTCCCCC-3'.

Additional primers which find use in the invention are deoxyribonucleic acid sequences having the formula:

$$5'-W_1-W_2-W_3-T-G-G-(T/G)-G-G-(C/G)-Z_1-3' \quad (IV)$$

wherein

W_1 is an oligonucleotide of from 7 to 12 bases selected from the group consisting of G-G-G-G-G-A-A-G-T-A-G-G, G-T-C-C-A-T-C-A-A, and A-G-C-G-A-G-G;

W_2 is an oligonucleotide of 2 bases selected from the group consisting of A-T and T-C;

W_3 is a base selected from the group consisting of T, G, and A; and

Z_1 is an oligonucleotide of from 3 to 7 bases selected from the group consisting of G-G-G, A-G-C-C-T-C-G and T-C-C;

with the provisos that at least approximately fifty percent (50%) of the deoxyribonucleotides of the

primer are deoxyguanylic acid and deoxycytidylic acid.

More usually, primers in accordance with formula IV will include sequences having the formula:

$$5'-W_1-A-T-W_3-T-G-G-T-(T/G)-G-G-(C/G)-Z_1-3' \qquad (V)$$

wherein W_1, W_3, and Z_1 are as previously defined. Presently preferred among primers of formula V are sequences having the formula:

$$5'-W_1-A-T-W_3-T-G-G-T-T-G-G-C-Z_1-3' \qquad (Va)$$

wherein W_1, W_3, and Z_1 are as previously defined. Of particular interest among the presently preferred primers of formula Va are those selected from the group consisting of:

5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'

5'-GTCCATCAATGTGGTTGGCAGCCTCG-3' and

5'-AGCGAGGATATGGTGGGCTCC-3'.

It will be recognized that primers other than those disclosed in the above formulae (I) to (Va) will be capable of hybridizing to at least a portion of the nucleic acid sequences in a nucleic acid sample and generating a plurality of amplification products therefrom in an amplification system. Thus, a person of ordinary skill in the art having the benefit of the present disclosure will readily find additional primers of use in the invention. For example, the following primers will provide additional exemplification of the practice of the

present invention:

5'-CAAGACCGGCAACAGG-3'
5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'
5'-GACCAACTGGTAATGGTAGCGACCGGC-3'
5'-ATGGCCTTCCAAAACGACGTCTA-3'
5'-CAAGACCGGCAACAGGATTC-3'
5'-TGGAGGAAGGGCTGGAGGAGGGCTCCGGAGGAAGGGC-3'
5'-GCCCTTCCTCCGGAGCCCTCCTCCAGCCCTTCCTCCA-3'
5'-GAGGTGGGCAGGTGGA-3'
5'-GACAGACAGACAGACA-3'
5'-TCCTAACCCTAAATCCAGCTCATGCC-3'
5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-CCCAACCAAGACCTACTTCC-3'
5'-ACTGACTGACTGACTG-3'
5'-TGTCTGTCTGTCTGTC-3'
5'-GCTCCGGAGGAAGGGC-3'
5'-GACACGACACGACACGACAC-3'
5'-CTCCTTCTCCAGCTGC-3'
5'-GAGGGTGGCGGTTCT-3'
5'-GAGGGTGGTGGCTCT-3'
5'-AGAACCGCCACCCTC-3'
5'-GGAGCTGGAGAAGGAG-3'
5'-TGGATGGATGGATGGATGGA-3'
5'-CGAGGCTGCCAACCACATTGATGAC-3'
5'-CACCACCACCACCAC-3'
5'-TTGCCTGTCTCCAGC-3'

Once the primer is hybridized to the template portion of nucleic acid sequences in the nucleic acid sample, a plurality of distinct amplification products can be produced by exposing the complexes to an amplification system which generates extension products from each hybridized primer.

A representative example of such an amplification system for a nucleic acid such as DNA will typically contain a pool of deoxyribonucleotides and a catalyst such as a DNA polymerase in a suitable buffer and under suitable reaction conditions (e.g., time, temperature, volume, etc.). For example, the deoxyribonucleoside triphosphates dATP, dCTP, dGTP, and dTTP will be added to the amplification system mixture in a buffered aqueous solution, preferably

adjusted to a pH of between 7 and 9, most preferably of approximately pH 8.

To this mixture will be added a molar excess of the selected primer. In the usual case, the amount of nucleic acid sequence available for hybridization with the primer will not be determinable with certainty. Thus, an excess of at least 1,000:1 (primer:sequence), and preferably an excess of at least 1,000,000:1 (primer:sequence), will generally be used for most nucleic acid samples. A large molar excess will generally be preferred in order to improve the efficiency of the amplification process.

To this mixture will be added an appropriate agent for catalyzing the primer extension reaction to produce amplification products. The catalyst selected for use in this invention may be any compound or system which will function to facilitate the synthesis of primer extension products. Suitable catalysts for this purpose include, for example, enzymes such as one or more of E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptases, and other enzymes, including heat stable enzymes which will facilitate combination of the nucleotides in the proper manner to form the primer extension products. A presently preferred catalyst is the DNA polymerase from Thermus aquaticus (Taq), as described in U.S. Patent No. 4,889,818, the entire disclosure of which is incorporated herein by this reference.

The temperature of the reaction mixture will range from, e.g., room temperature up to a temperature above which the catalyst no longer functions efficiently. The selected temperature will depend in part on the particular catalyst used. For example, some DNA polymerases are used at a

temperature generally no greater than about 40°C (e.g. Klenow). Other DNA polymerases can be used at much higher temperatures (e.g., Taq polymerase is generally used at 72°C).

Generally, the synthesis will be initiated at the 5'-end of each primer and proceed in the 3'-direction along the nucleic acid sequence template strand until synthesis terminates. However, there may be catalysts which initiate synthesis at the 3'-end and proceed in the other direction using the process substantially as described above.

In evaluating the present invention, it has been found useful to compare the novel compositions and methods of the present invention with a known amplification procedure, albeit one which is not designed to detect and analyze genomic variation. As noted above, the PCR procedure uses two primers with the intent of amplifying a single sequence whose length is defined by the position of the primers. By contrast, it is the object of the present invention to amplify a plurality of sequences in a nucleic acid sample, each having a different length (and consequently a different molecular weight). In certain preferred embodiments, many such sequences are amplified so that genomic variations in the sequences may be identified and correlated to genetically related individuals or groups.

This difference can be well-illustrated by using known primers and known template sequence. While this knowledge is useful for understanding certain aspects of the present invention, the practice of the invention does not require that the template sequences be known in this manner or that the primers be constructed with

TABLE 1 - CHICKEN GLOBIN SEQUENCE

1	GCTCCACCTG	CACCAAGGCA	GACCCTAACC	TCAACCCATG	CCGGTGGAGA	CACTAACCCT
61	AATCCTAACC	CTAAATCCAG	CTCATGCCAG	CATCACACTG	CCCCAACCCT	AACACCAAGC
121	TCAACCCTGA	TCCTAACACT	AACCCCAGCT	CGCGTCGGGG	TCCAACCCCC	CCAGCCTGCG
181	CAGTATCGTG	GGTGGCGAGG	GCAGCAGCCC	TGCCTGGCTG	GGGTCCAGAA	TCTATGGGGG
241	GGGCTGGGGG	GTGGGCGGTG	GCCAGCACAG	CATATAAGGC	TGACAGCAGA	CTTCAGGGGC
301	ACCCGTGCTG	GGGGCTGCCA	ACACAGAGGT	GCAACCATGG	TGCTGTCCGC	TGCTGACAAG
361	AACAACGTCA	AGGGCATCTT	CACCAAAATC	GCCGGCCATG	CTGAGGAGTA	TGGCGCCGAG
421	ACCCTGGAAA	GGTAGGTGTC	CTTCTCTGCC	TCCGGCTGCC	TCTCTCCCCT	GATCCCCTYC
481	CCGTCCTCAG	CTGCCCCCGT	CCTATCCCTC	CCTGCCTTAC	CCGTCCCTCT	CCCCTCCTCC
541	CCTGACTCAC	TGTGCTCCGC	AGGATGTTCA	CCACCTACCC	CCCAACCAAG	ACCTACTTCC
601	CCCACTTCGA	TCTGTCACAC	GGCTCCGCTC	AGATCAAGGG	GCACGGCAAG	AAGGTAGTGG
661	CTGCCTTGAT	CGAGGCTGCC	AACCACATTG	ATGACATCGC	CGGCACCCTC	TCCAAGCTCA
721	GCGACCTCCA	TGCCCACAAG	CTCCGCGTGG	ACCCTGTCAA	CTTCAAAGTG	AGTGTCYGGG
781	AAGGGGCGAC	CCGCCGCCCC	CACCGCGTGG	ACCCTGTCAA	CTTCAAAGTG	AGTGTCYGGG
841	TCACCCCCTC	GCTCATCCTC	TCCTTTTGCC	TTGCAGCTCC	TGGGCCAATG	CTTCCTGGTG
901	GTGGTGGCCA	TCCACCACCC	TGCTGCCCTG	ACCCCGGAGG	TCCATGCTTC	CCTGGACAAG
961	TTCTTGTGCG	CCGTGGGCAC	TGTGCCCTG	ACCCCGGAGG	TCCATGCTTC	CCTGGACAAG
1021	AGAGCTGGGG	CCAACCCATC	GCCAGCCCTC	CGACAGCGAG	CAGCCAAATG	AGATGAAATA
1081	AAATCTGTTG	CATTTGTGCT	CCAGCCCTGG	TGTTCTGCTT	TGGGGCTCTG	TGCAGCTGAA
1141	CCAGGGCCCC	CGGGGCCAGG	TGGGTGTGCG	GAGGATATGG	GCTCAGGTGG	TGGTGTTTGA
1201	GTGGTGTGCC	CAGAAG				

this knowledge. However, for illustrative purposes, primers from the chicken α-globin sequence can be employed. The sequence [from J.B. Dodgson and J.D. Engel, J. Bio. Chem. 258:4623 (1983)] is provided in Table 1.

Primer 226A corresponds to positions 64-88 in the 5'-end flanking region of the α-globin sequence. Primer 227 corresponds to positions 578-603 (opposite strand) in the Exon 2 region of the α-globin sequence. In the PCR, with a nucleic acid source containing the α-globin sequence, primers 226 and 227 can be used for amplification of a 540 bp product.

In the present comparitive illustration, primers 226 and 227 were used, both separately and together, in nine separate amplification reactions (Table 2). The products of the reactions were electrophoresed and a photograph of the ethidium bromide stained gel is shown as Figure 1A. Molecular weight markers were provided in Lanes 1 and 12. Lanes 2-10 correspond to the nine amplification reactions (see Table 2).

Table 2

Amplification of Chicken Genomic DNA: PCR versus SPAR

Lane[a]	Primer Concentration[b] 226	227	226:227 Primer Ratio
2	0.0	0.5	-
3	0.5	0.5	1:1
4	0.5	0.25	2:1
5	0.5	0.120	4:1
6	0.5	0.050	10:1
7	0.5	0.025	20:1
8	0.5	0.005	100:1
9	0.5	0.0025	200:1
10	0.5	0.0	-

[a] see Figure 1A

[b] given in micrograms and assuming the same reaction volume

Where primer 227 was used alone (lane 2), no amplification bands are observed. Where primers 226 and 227 are used in equal amounts (lane 3), a single PCR-product band is observed running at approximately

the expected molecular weight for a 540 bp product. Where the ratio of primer 226 to primer 227 is 2:1 (lane 4) or 4:1 (lane 5), the same single product band is observed but with some increased background. On the other hand, where the ratio is 10:1 (lane 6) or greater (lanes 7-9), a host of new product bands are apparent. Surprisingly, the new product bands can be generated using primer 226 alone (lane 10). Interestingly, the single product band observed clearly in lanes 3-5 becomes less apparent as the concentration of primer 227 decreases until it disappears (lane 10).

To examine the nature of the product bands of the amplifications, the similarity of the single product band (lanes 3-5, Figure 1A) with the multiple product bands (lanes 6-10, FIgure 1A) was determined. For this determination, the products of the gel (Figure 1A) were transferred to nitrocellulose (Schleicher and Schuell, NH; BA-S 85 nitrocellulose membrane) by blotting. Thereafter, primer 267 (5'-ACACAGAGGTGCAACC-3', corresponding to positions 320-336 in the 5'-end flanking region of the α-globin sequence (see Table 1), a region internal to the region defined by the annealing boundaries of primers 226 and 227) was radiolabelled and used as a probe on the nitrocellulose. Following a period that would allow hybridization, the nitrocellulose was washed and autoradiographed (Figure 1B). Strong signal is observed in what corresponds to lanes 3-7 of Figure 1A. A faint signal is observed in lanes 8 and 9, while no signal is observed in lane 10. Measured distances on the original electrophoresed gel (Figure 1A) confirm that the signal produced represents the single PCR-product band resulting from a two primer reaction. Primer 267 does not hybridize with any of the multiple band products, indicating that these

bands do not appear to share this sequence with the internal 489 bp region of the 540 bp product of PCR amplification.

From the above comparative illustration it is clear that the present invention, using as few as one primer for amplification, i.e. a Single Primer Amplification Reaction ("SPAR"), amplifies a plurality of sequences found in a nucleic acid sample. The primer is used in conjunction with a nucleic acid sample comprising nucleic acid sequences of a distinct nucleic acid source. The selected primer will be capable of hybridizing to at least a portion of the nucleic acid sequences. However, it is not necessary that the sequences are known.

In certain preferred embodiments, the primer comprises an oligonucleotide capable of hybridizing to a plurality of regions (templates) within the nucleic acid sequences, under conditions which the primer serves as the initiating sequence for the synthesis of an extension product which is complementary to a portion of the nucleic acid sequences, and wherein the extension product, when separated from the nucleic acid sequence, is also capable of hybridizing to the primer. In this latter case, the primer contemplated by the present invention will detect inverted repeat nucleic acid sequences in the nucleic acid sample.

In determining the number of extension products to be expected in an amplification of a given nucleic acid sample, a number of considerations must be addressed. Inverted repeats are known to exist throughout a genome. W.R. Jelinek and C.W. Schmid, Ann. Rev. Biochem. 51:813 (1982). However, conventional PCR requires that the primers hybridize within approximately 3kb of each other (the upper size limit). It is not currently known that inverted

repeats exist within this constraint in a given genome in any frequency. However, estimations regarding the probability of such occurrence can be made.

For example, in a sequence of four billion bases (approximately the size of the human genome), there are about four billion n-mers for a given n, where an n-mer is an oligodeoxynucleotide of length n bases. The random probability of the occurrence of a particular n-mer with a specific sequence, assuming equal probability of all four bases, is: $(1/4)^n$. The random probability of a particular n-mer among 4 billion n-mers is thus:

$$(4 \times 10^9)(1/4)^n$$

The probability of a second n-mer appearing within 3kb (the PCR upper size limit) is:

$$(3 \times 10^3)(1/4)^n$$

The probability of such an occurrence for each n-mer class is then:

$$(1.2 \times 10^{13})((1/4)^n)^2$$

Therefore, the minimum length of a primer which will have a non-random inverted repeat within a given genome can be calculated, where the probability is less than one. A compilation of the approximate probabilities for selected n-mers in a genome of 4×10^9 is as follows:

n-mer	Probability
10	10.9
11	5.7×10^{-1}
12	3.5×10^{-2}
16	5.4×10^{-7}
32	2.9×10^{-26}

Since $(1/4)^{11} = 2.4 \times 10^{-7}$, the expectation is for any 11-mer to appear by chance 960 times in a genome of 4×10^9 bases. Therefore, the probability of two 11-mers chosen at random to occur within 3kb of each other, hence among 3,000 11-mers, with their 3'-ends pointing towards each other, is:

$(3.0 \times 10^3)(4 \times 10^9)(2.4 \times 10^{-7})^2 = 5.7 \times 10^{-1}$

At this probability, one would expect only infrequently to produce multiple amplified extension products. Similarly, if the requirement is for precisely identical 11-mers, the probability of two with a specified sequence appearing by chance within 3kb of each other, and with the 3'-ends pointing towards each other, is less than one in a genome of four billion bases.

Whether perfect matches are required in SPAR was investigated indirectly. Having determined that a single primer can be used to generate multiple bands in an amplification system (Figure 1A), variants of this particular primer were synthesized, most differing from primer 226A by only one nucleotide (Table 3).

Table 3

Primer Complementarity

Primer	Lane[a]	Sequence[b]
226A	2,12	5'-TCCTAACCCTAAATCCAGCTCATGCC-3'
226	3,13	5'-C<u>C</u>CTAACCCTAAATCCAGCTCATGCC-3'
268	4,14	5'-TCCTAACCCTAAATCCAGCTCAT<u>C</u>CC-3'
269	5,15	5'-TCCTAACCCTAAATCCAGCTCAT<u>A</u>CC-3'
270	6,16	5'-TC<u>G</u>TAACCCTAAATCCAGCTCATGCC-3'
271	7,17	5'-TC<u>A</u>TAACCCTAAATCCAGCTCATGCC-3'
272	8,18	5'-TCCTAACCCTAAAT<u>G</u>CAGCTCATGCC-3'
273	9,19	5'-TCCTAACCCTAAAT<u>A</u>CAGCTCATGCC-3'
246	10,20	5'-ACCCTAAATCCAGCTC-3'

[a] lanes are with reference to Figure 2 (45°C,55°C)

[b] nucleotide substitutions are underlined with reference to primer 226A

Note that primer 246, a 16-mer, is completely homologous to an internal sequence of primer 226A, but lacks five nucleotides on both the 3'-end and 5'-end. SPAR was carried out as in Figure 1, except that annealing was performed at two temperatures (45°C and 55°C) for two (2) minutes. SPAR cycles were carried out in this manner for 35 cycles followed by 15 minutes at 70°C in the final cycle. The products were electrophoresed, the gel and the stained gel photographed (Figure 2). Molecular weight markers are in lanes 1, 11 and 21.

Figure 2 shows that primers differing by a single nucleotide may produce a multiple band pattern which is almost completely different (e.g. compare lanes 2 and 4), while in other cases, primers that differ by a single nucleotide will produce almost the same multiple band patterns (e.g., compare lanes 6 and 7).

Figure 2 also shows that the annealing temperature can change the multiple band pattern generated by SPAR (e.g. compare lanes 3 and 13). To illustrate further the impact of annealing temperature on the multiple band pattern generated by SPAR, primers 226A and 235 were used. Primer 235 corresponds to positions 671-695 (opposite strand) of the α-globin sequence (see Table 1). In the PCR, with a nucleic acid source containing the α-globin sequence, primers 226A and 235 can be used together to define a 581 bp region between their annealing boundaries and allow for amplification of a 632 bp product.

In this illustration, the ratio of the primers was varied in the manner described earlier. The primer ratios for the amplifications are shown in Table 4.

Table 4

Amplification of Chicken Genomic DNA: PCR versus SPAR

Lane[a]	Primer Concentration[b] 226A	235	226A:235 Primer Ratio
2	0.5	0.5	1:1
3	0.5	0.25	2:1
4	0.5	0.120	4:1
5	0.5	0.050	10:1
6	0.5	0.025	20:1
7	0.5	0.005	100:1
8	0.5	0.0025	200:1
9	0.5	0.0	-
10	0.0	0.5	-

[a] see Figures 3A, 3B and 3C

[b] given in micrograms and assuming the same reaction volume

Three annealing temperatures were compared: 55°C (Figure 3A), 60°C (Figure 3B) and 65°C (Figure 3C).

Figure 3A shows the expected single PCR-product band (arrow) as well as a multiple band pattern. As was seen in Figure 1A, the single PCR-product band in Figure 3A gradually decreases in intensity (see lanes 2-8) and disappears when primer 226A is used alone (lane 9). New bands, however, appear as the ratio of 226A primer to 235 primer increases.

Figure 3B shows the results when the same amplifications are carried out using a 60°C annealing temperature. Again the single product band expected for PCR (arrow) decreases in intensity as the ratio of 226A primer to 235 primer increases (see lanes 2-7). Similarly, a multiband pattern (less complex than observed in Figure 3A) begins to develop as this ratio changes (see in particular lanes 5 and 6). Very little signal, however, is apparent at this annealing temperature when primer 226A is used alone in a SPAR (see lane 9).

Finally, Figure 3C shows the results when the same amplifications are carried out using a 65°C

annealing temperature. Again the single product band expected for PCR (arrow) decreases in intensity as the ratio of 226A primer to 235 primer increases (see lanes 2-7). A weak multiband pattern is evident, although the bands are not intense as this ratio changes (see in particular lanes 5 and 6). However, no signal is apparent at this annealing temperature when primer 226A is used alone in a SPAR reaction (see lane 9).

From the above it is clear that single primers can be used to amplify multiple sequences in a SPAR. While not limited to any theory, it is believed that the primers capable of producing amplification products according to the present invention recognize template sequences which are not randomly distributed throughout the genome. The primer is believed to permit the synthesis of an extension product which is complementary to a portion of the nucleic acid sequences, and wherein the extension product, when separated from the nucleic acid sequence, is also capable of hybridizing to the primer. Regardless of the mechanism, the multiple sequences generated by SPAR provide a rich source of markers for detection and analysis of genomic variation.

Certain embodiments of the present invention contemplate comparisons between the nucleic acid extension products from at least two sources (e.g. different individuals within a species, individuals from different species, etc.). It is desired that the amplified nucleic acid sequences from each nucleic acid source be thereafter compared to determine the degree of homology of the nucleic acid sequences.

In certain preferred embodiments, the degree of "size homology" of the amplification products is measured, e.g., the amplified nucleic acid sequences

from each nucleic acid source are electrophoresed side-by-side in a gel. The various amplification products are, in this manner, separated by size, appearing as bands in the gel. Where there are a plurality of amplification products from one nucleic acid source, a distinct migration pattern of bands will be evident. The different migration patterns of bands evident between two sources of nucleic acid is a measure of the degree of "size homology" because different size products indicate different sequences.

The present invention contemplates measuring the degree of homology of the amplified nucleic acid sequences by other methods as well. While not required, the present invention also contemplates measuring the degree of "sequence homology" of the amplified nucleic acid sequences by a) hybridization, b) restriction digestion and/or c) sequencing. If hybridization is used, the size-separated amplification products are transferred to a suitable blotting medium (e.g., nitrocellulose) and hybridization can be carried out in the manner of Southern et al. described above. Where restriction digestion is used, the amplified nucleic acid sample is digested with one or more restriction enzymes prior to or after electrophoresis. Where sequencing is used, the individual nucleic acid bands of the gel are recovered and sequenced.

With the present invention, it is possible to amplify nucleic acid in genomic DNA to a level detectable by several different methodologies (e.g. hybridization with a labelled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labelled CTP or ATP into the amplified segment). However, it is generally preferred that amplification proceed to allow chemical amounts of amplified nucleic acid to

be created. This allows for detection by simple means such as ethidium bromide staining.

EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μM (micromolar); N (Normal); mol (moles); mmol (millimoles); μmol (micromoles); nmol (nanomoles); gm (grams); mg (milligrams); μg (micrograms); L (liters); ml (milliliters); μl (microliters); cm (centimeters); mm (millimeters); μm (micrometers); nm (nanometers); °C (degrees Centigrade); MW (molecular weight); OD (optical density); EDTA (ethylenediaminetetraacetic acid); dNTP (deoxyribonucleoside 5'-triphosphate); TAE buffer (0.06 M Tris-acetate, pH 8.3; 0.003 M EDTA); Taq buffer (50 mM KCl, 2.5 mM $MgCl_2$, 10 mM Tris, pH 8.5, 200 μg/ml gelatin); PAGE (polyacrylamide gel electrophoresis); V (volts); W (watts); mA (milliamps); bp (base pair); CPM (counts per minute).

Generally, presently preferred embodiments of the present method will be carried out using approximately 8 μl dNTPs (each at 2.5 μM, totaling 10μM) and 100 ng to 1 μg of a selected primer. As a catalyst, Taq polymerase (approximately 0.5 μl; 5 Units/μl, Bethesda Research Laboratories, Gaithersburg, MD) will be used and the reaction will ordinarily be performed in a total volume of approximately 50-100 μl made up with buffer containing, e.g. 10 mM Tris, pH 8.5, 50 mM KCl, 1-2.5 mM $MgCl_2$, and 100-200 μg/ml gelatin. The SPAR will

typically be performed in a programmable thermal cycler (ERICOMP, Inc., San Diego, CA); a small amount of mineral oil or paraffin oil in each well can be used for maximum efficiency of heat exchange. In a typical SPAR amplification cycle, denaturation will be at 95°C for one minute; annealing at 55°C for two minutes; and extension at 70°C for three to eight minutes. SPAR cycles will normally be carried out in this manner for 25-35 cycles (approximately 12 minutes per cycle) followed by approximately 15 minutes at 70°C.

Where gel electrophoresis is used, 200-300 ml gels of 1.8% SeaKem® agarose (FMC BioProducts, Rockland, Maine) will be poured and then electrophoresed for approximately 2-7 hours (100V, 130mA) in TAE buffer. Following electrophoresis, individual bands will generally be visualized by ethidium bromide staining (approximately 30 minutes) followed by a destaining wash in water.

Selected primers used in practice of the present invention are listed in Table 5. In addition to providing the sequence and origin of each primer, Table 5 provides reference numbers used to identify the particular primers. For example, primer 127 is derived from the λgt11 β-galactosidase gene near the EcoRI site. T.V. Huynh et al. describe a lambda vector λgt11 [In: DNA Cloning: A practical approach, IRL Press, 1:49 (1985)], which carries a portion of the E. coli β-galactosidase gene, including the upstream elements. There is, within the carboxy-a

SUBSTITUTE SHEET

TABLE 5 - OLIGONUCLEOTIDE SEQUENCES

No.	N-mer	%GC	Origin	Sequence
89	16	57	3'-end of Nopaline synthase gene from A. tumefaciens	5'-CAAGACCGGCAACAGG-3'
95	33	55	3'-end of Maize Lipid Body 3 Protein gene	5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'
127	27	60	λgt11 β-galactosidase gene near EcoRI site	5'-GACCAACTGGTAATGGTAGCGACCGGC-3'
162	23	48	rat delta-9 desaturase gene	5'-ATGGCCTTCCAAAACGACGTCTA-3'
170	20	55	3'-end of Nopaline synthase gene from A. tumefaciens	5'-CAAGACCGGCAACAGGATTC-3'
201	37	71	Consensus repeat in Jeffreys' human 33.6 probe	5'-TGGAGGAAGGGCTGGAGGAGGGCTCCGGAGGAAGGGC-3'
201'	37	71	Complement of primer 201	5'-GCCCTTCCTCCGGAGCCCTCCTCCAGCCCTTCCTCCA-3'
202	16	69	Consensus repeat in Jeffreys' human 33.15 probe	5'-GAGGTGGGCAGGTGGA-3'
216	16	50	Synthetic (probes multibands in barley)	5'-GACAGACAGACAGACA-3'
226A	26	50	Chicken α-globin gene	5'-TCCTAACCCTAAATCCAGCTCATGCC-3'
227	26	66	Chicken α-globin gene	5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'
235	25	56	Chicken α-globin gene	5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
239	21	62	Chicken cytochrome b5 gene	5'-AGCGAGGATATGGTGGGCTCC-3'
242	20	55	Chicken α-globin. Complement of 227	5'-CCCAACCAAGACCTACTTCC-3'
243	16	50	Synthetic	5'-ACTGACTGACTGACTG-3'
244	16	50	Synthetic	5'-TGTCTGTCTGTCTGTC-3'
245	16	75	Subset of Jeffreys' 37 mer	5'-GCTCCGGAGGAAGGGC-3'
247	20	60	Synthetic	5'-GACACGACACGACACGACAC-3'
250	16	63	Homeo box consensus sequent	5'-CTCCTTCTCCAGCTGC-3'
251	15	67	M13 repeat unit	5'-GAGGGTGGCGGTTCT-3'
252	15	67	M13 repeat unit	5'-GAGGGTGGTGGCTCT-3'
253	15	67	M13 repeat unit	5'-AGAACCGCCACCCTC-3'
258	16	63	Complement of 250	5'-GGAGCTGGAGAAGGAG-3'
259	20	50	5' to human myelin basic protein	5'-TGGATGGATGGATGGATGGA-3'
260	25	56	Complement of 235	5'-CGAGGCTGCCAACCACATTGATGAC-3'
261	15	67	Synthetic	5'-CACCACCACCACCAC-3'
277	15	60	Part of a repeat unit in cattle	5'-TTGCCTGTCTCCAGC-3'

terminal coding region of E. coli β-galactosidase gene, a single EcoRI site into which foreign DNA can be inserted. The λgt series of insertion vectors was designed to express cDNA as β-galactosidase fusion protein; DNA fragments (up to 7.2 kb) are cloned into the unique EcoRI site located in lacZ, allowing expression of a fusion protein if the cloned sequence is properly in-frame.

Routinely, the primers (see Table 5) will be synthesized, for example on a CYCLONE™ DNA Synthesizer (BIOSEARCH, Inc., San Rafael, CA). This instrument automates solid phase, phosphoramidite synthesis of DNA fragments on derivatized controlled pore glass supports.

EXAMPLE 1

In this example, the multiple band pattern generated by SPAR is investigated with different sources of genomic nucleic acid. Two different primers are used (226 and 216) separately in each amplification with each nucleic acid source. Primer 226 (see Table 5) has been fully described above. Primer 216, a 16-mer, was synthesized having the repeating unit -GACA- (see Table 5).

Genomic DNA was extracted from eight individual chickens and used to provide nucleic acid samples in SPAR. The amplified products were evaluated by electrophoresis. Figure 4 is a photograph of the ethidium bromide stained gel. Lane I contains molecular weight markers. Lanes 2-9 show that the amplification products for the individual chicken nucleic acid samples using primer 216 comprise a multi-band pattern in every case. While a number of prominent bands are seen to be shared, primer 216 also produces amplification product bands associated with only a few of the individual chickens (arrows),

indicating polymorphisms. Lane 10 is a control lane (no nucleic acid sample). Lanes 11-17 show that the amplification products for the nucleic acid samples using primer 226 also comprise a multiband pattern in every case. Again, certain prominent bands found in each pattern, while other bands are associated only with particular individuals (arrows), indicating polymorphisms.

EXAMPLE 2

In this example, multiple band patterns generated by SPAR are produced with different sources of genomic nucleic acid. Primer 89, a 16-mer (see Table 5), was used in each amplification with each nucleic acid source.

Genomic DNA was extracted from twenty-four (24) different corn plants, each from a different inbred corn line, and used as samples in SPAR. Amplification was carried out using 10 µl Taq buffer (10X), 0.3 µl of each dNTP (100 mM), 2 µl gelatin (10 mg/ml), 0.5 µl Taq polymerase and 1 µl of spermidine (100 mM) in a total volume of 100 µl (brought up in water). Denaturation was at 94°C for one (1) minute; annealing was at 48°C for two (2) minutes; and extension at 72°C for three (3) minutes. The SPAR was carried out in this manner for 24 cycles, then extension was carried out for ten (10) minutes in a final cycle. The products were electrophoresed as described above, the gel stained, and the stained gel photographed (Figure 5A and 5B). Molecular weight markers were provided in the end lanes.

Lanes 2-14 of Figure 5A show that the amplification products for the individual corn lines using primer 89 comprise a multi-band pattern in all but one case. While a number of prominent bands are found in each pattern, primer 89 also produces

amplification product bands associated with only a few of the individual corn lines (arrows), indicating polymorphisms. Similarly, lanes 2-12 of Figure 5B show that the amplification products for the individual corn lines using primer 89 comprise a multiband pattern (this time in every case). Again, certain strong bands are shared, while other bands may be associated only with particular individuals (arrows), indicating polymorphisms.

From the above it is evident that the present invention provides a method offering information as important and useful as RFLP data without the accompanying labor, time and expense. The present invention does not require the use of restriction enzymes or gel transfer. The present invention provides primers as well as a single primer amplification method for amplification of multiple sequences in chemical amounts using template representing the entire genome. The multiple sequences provide a rich source of markers for analysis of genomic variation.

All patent publications cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference.

Claims

1. A method for amplifying a plurality of sequences found in a nucleic acid sample comprising:

a) providing a nucleic acid sample comprising nucleic acid sequences of a distinct nucleic acid source;

b) providing a single oligonucleotide primer comprising a sequence of at least eleven nucleotides which is capable of hybridizing to at least a portion of said nucleic acid sequences and generating a plurality of amplification products therefrom in an amplification system; and

c) bringing together said nucleic acid sample with said primer in an amplification system, thereby amplifying at least one portion of said nucleic acid sequences, so that discrete portions of said nucleic acid sequences are detectable.

2. The method of claim 1 wherein said nucleic acid sequences represent the entire genome of said distinct nucleic acid source.

3. The method of claim 1 wherein said distinct nucleic acid source is selected from the group consisting of virus, bacteria, fungi, yeast, algae, mycoplasma and protozoa.

4. The method of claim 1 wherein said distinct nucleic acid source is an animal selected from the group consisting of mammals, birds, reptiles, amphibians and insects.

5. The method of claim 4 wherein said animal nucleic acid source is selected from the group consisting of human, horse, dog, cow, salmon, chicken, mouse, rat and bee.

6. The method of claim 1 wherein said distinct nucleic acid source is a plant selected from the group consisting of angiosperms and gymnosperms.

7. The method of claim 6 wherein said plant nucleic acid source is selected from the group consisting of corn, barley, wheat, apple, alfalfa, soybean, oil rape, tobacco, tomato and loblolly pine.

8. The method of claim 1 wherein said primer has a sequence of deoxyribonucleotides of the formula:

$$5'-X_1-X_2-X_3-X_4-G-A-C-Y_1-Y_2-Y_3-3'$$

wherein

X_1 is desX_1 or an oligonucleotide of from 3 to 11 bases selected from the group consisting of A-G-A-G, G-A-C-C-A-A-C-T-G-G-T, and C-C-C;

X_2 is desX_2 or an oligonucleotide of 3 bases selected from the group consisting of A-A-T and C-C-C;

X_3 is desX_3 or an oligonucleotide of from 1 to 3 bases selected from the group consisting of T-C-T, G-G-T, A-A-C, and G;

X_4 is an oligonucleotide of 3 bases selected from the group consisting of C-A-A, A-G-C and A-C-A;

Y_1 is a base selected from the group consisting of C and A;

Y_2 is an oligonucleotide of 3 bases selected from the group consisting of G-G-C, C-C-T, G-A-C, and T-A-C; and

Y_3 is desX_3 or an oligonucleotide of from 5 to 13 bases selected from the group consisting of A-A-C-A-G-G, G-G-G-C-C-T-G-G-T-C-G-A-T, A-G-A-C-A and T-T-C-C-C-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

9. The method of claim 8 wherein said sequence has the formula:

$5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-Y_2-Y_3-3'$

wherein X_1, X_2, X_3, Y_2 and Y_3 are as previously defined.

10. The method of claim 8 wherein said sequence has the formula:

$5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-G-(G/A)-C-Y_3-3'$

wherein X_1, X_2, X_3 and Y_3 are as previously defined.

11. The method of claim 8 wherein said sequence is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'

5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'

5'-GACCAACTGGTAATGGTAGCGACCGGC-3'

5'-GACAGACAGACAGACA-3' and

5'-CCCCCCAACCAAGACCTACTTCCCCC-3'.

12. The method of claim 1 wherein said primer has a sequence of deoxyribonucleotides of the formula:

$5'-W_1-W_2-W_3-T-G-G-(T/G)-G-G-(C/G)-Z_1-3'$

wherein

W_1 is an oligonucleotide of from 7 to 12 bases selected from the group consisting of G-G-G-G-G-A-A-G-T-A-G-G, G-T-C-C-A-T-C-A-A, and A-G-C-G-A-G-G;

W_2 is an oligonucleotide of 2 bases selected from the group consisting of A-T and T-C;

W_3 is a base selected from the group consisting of T, G, and A; and

Z_1 is an oligonucleotide of from 3 to 7 bases selected from the group consisting of G-G-G, A-G-C-C-T-C-G and T-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

13. The method of claim 12 wherein said sequence has the formula:

5'-W_1-A-T-W_3-T-G-G-T-(T/G)-G-G-(C/G)-Z_1-3'

wherein W_1, W_3, and Z_1 are as previously defined.

14. The method of claim 12 wherein said sequence has the formula:

5'-W_1-A-T-W_3-T-G-G-T-T-C-G-C-Z_1-3'

wherein W_1, W_3, and Z_1 are as previously defined.

15. The method of claim 12 wherein said sequence is selected from the group consisting of:

5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'

5'-GTCCATCAATGTGGTTGGCAGCCTCG-3' and

5'-AGCGAGGATATGGTGGGCTCC-3'.

16. The method of claim 1 wherein said primer is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'
5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'
5'-GACCAACTGGTAATGGTAGCGACCGGC-3'
5'-ATGGCCTTCCAAAACGACGTCTA-3'
5'-CAAGACCGGCAACAGGATTC-3'
5'-TGGAGGAAGGGCTGGAGGAGGGCTCCGGAGGAAGGGC-3'
5'-GCCCTTCCTCCGGAGCCCTCCTCCAGCCCTTCCTCCA-3'
5'-GAGGTGGGCAGGTGGA-3'
5'-GACAGACAGACAGACA-3'
5'-TCCTAACCCTAAATCCAGCTCATGCC-3'

5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-CCCAACCAAGACCTACTTCC-3'
5'-ACTGACTGACTGACTG-3'
5'-TGTCTGTCTGTCTGTC-3'
5'-GCTCCGGAGGAAGGGC-3'
5'-GACACGACACGACACGACAC-3'
5'-CTCCTTCTCCAGCTGC-3'
5'-GAGGGTGGCGGTTCT-3'
5'-GAGGGTGGTGGCTCT-3'
5'-AGAACCGCCACCCTC-3'
5'-GGAGCTGGAGAAGGAG-3'
5'-TGGATGGATGGATGGATGGA-3'
5'-CGAGGCTGCCAACCACATTGATGAC-3'
5'-CACCACCACCACCAC-3'
5'-TTGCCTGTCTCCAGC-3'.

17. A method for detecting variation between nucleic acid samples, comprising:

a) providing at least two nucleic acid samples each comprising nucleic acid sequences representing the entire genome of a distinct nucleic acid source;

b) providing a single oligonucleotide primer comprising a sequence having at least eleven nucleotides and capable of hybridizing to at least a portion of said nucleic acid sequences from each source and generating a plurality of amplification products therefrom in an amplification system; and

c) bringing together each nucleic acid sample with said primer in a separate amplification system, thereby amplifying at least one portion of said nucleic acid sequences from at least one source, so that discrete portions of said nucleic acid sequences are detectable.

18. A method as recited in claim 17, further comprising:

d) comparing the amplified portions of nucleic acid sequences from each nucleic acid source so amplified to determine the degree of homology between said nucleic acid sources.

19. The method of claim 17 wherein said nucleic acid sequences represent the entire genome of said distinct nucleic acid source.

20. The method of claim 17 wherein said distinct nucleic acid source is selected from the group consisting of virus, bacteria, fungi, yeast, algae, mycoplasma and protozoa.

21. The method of claim 17 wherein said distinct nucleic acid source is an animal selected from the group consisting of mammals, birds, reptiles, amphibians and insects.

22. The method of claim 21 wherein said animal nucleic acid source is selected from the group consisting of human, horse, dog, cow, salmon, chicken, mouse, rat and bee.

23. The method of claim 17 wherein said distinct nucleic acid source is a plant selected from the group consisting of angiosperms and gymnosperms.

24. The method of claim 23 wherein said plant nucleic acid source is selected from the group consisting of corn, barley, wheat, apple, alfalfa, soybean, oil rape, tobacco, tomato and loblolly pine.

25. The method of claim 17 wherein said primer has a sequence of deoxyribonucleotides of the formula:

$$5'-X_1-X_2-X_3-X_4-G-A-C-Y_1-Y_2-Y_3-3'$$

wherein

X_1 is desX_1 or an oligonucleotide of from 3 to 11 bases selected from the group consisting of A-G-A-G, G-A-C-C-A-A-C-T-G-G-T, and C-C-C;

X_2 is desX_2 or an oligonucleotide of 3 bases selected from the group consisting of A-A-T and C-C-C;

X_3 is desX_3 or an oligonucleotide of from 1 to 3 bases selected from the group consisting of T-C-T, G-G-T, A-A-C, and G;

X_4 is an oligonucleotide of 3 bases selected from the group consisting of C-A-A, A-G-C and A-C-A;

Y_1 is a base selected from the group consisting of C and A;

Y_2 is an oligonucleotide of 3 bases selected from the group consisting of G-G-C, C-C-T, G-A-C, and T-A-C; and

Y_3 is desX_3 or an oligonucleotide of from 5 to 13 bases selected from the group consisting of A-A-C-A-G-G, G-G-G-C-C-T-G-G-T-C-G-A-T, A-G-A-C-A and T-T-C-C-C-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

26. The method of claim 25 wherein said sequence has the formula:

$$5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-Y_2-Y_3-3'$$

wherein X_1, X_2, X_3, Y_2 and Y_3 are as previously defined.

27. The method of claim 25 wherein said sequence has the formula:

5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-G-(G/A)-C-Y_3-3'

wherein X_1, X_2, X_3 and Y_3 are as previously defined.

28. The method of claim 25 wherein said sequence is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'

5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'

5'-GACCAACTGGTAATGGTAGCGACCGGC-3'

5'-GACAGACAGACAGACA-3' and

5'-CCCCCCAACCAAGACCTACTTCCCCC-3'.

29. The method of claim 17 wherein said primer has a sequence of deoxyribonucleotides of the formula:

5'-W_1-W_2-W_3-T-G-G-(T/G)-G-G-(C/G)-Z_1-3'

wherein

W_1 is an oligonucleotide of from 7 to 12 bases selected from the group consisting of G-G-G-G-A-A-G-T-A-G-G, G-T-C-C-A-T-C-A-A, and A-G-C-G-A-G-G;

W_2 is an oligonucleotide of 2 bases selected from the group consisting of A-T and T-C;

W_3 is a base selected from the group consisting of T, G, and A; and

Z_1 is an oligonucleotide of from 3 to 7 bases selected from the group consisting of G-G-G, A-G-C-C-T-C-G and T-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

30. The method of claim 29 wherein said sequence has the formula:

5'-W_1-A-T-W_2-T-G-G-T-(T/G)-G-G-(C/G)-Z_1-3'

wherein W_1, W_3, and Z_1 are as previously defined.

31. The method of claim 29 wherein said sequence has the formula:

5'-W_1-A-T-W_3-T-G-G-T-T-G-G-C-Z_1-3'

wherein W_1, W_3, and Z_1 are as previously defined.

32. The method of claim 29 wherein said sequence is selected from the group consisting of:

5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'

5'-GTCCATCAATGTGGTTGGCAGCCTCG-3' and

5'-AGCGAGGATATGGTGGGCTCC-3'.

33. The method of claim 17 wherein said primer is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'
5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'
5'-GACCAACTGGTAATGGTAGCGACCGGC-3'
5'-ATGGCCTTCCAAAACGACGTCTA-3'
5'-CAAGACCGGCAACAGGATTC-3'
5'-TGGAGGAAGGGCTGGAGGAGGGCTCCGGAGGAAGGGC-3'
5'-GCCCTTCCTCCGGAGCCCTCCTCCAGCCCTTCCTCCA-3'
5'-GAGGTGGGCAGGTGGA-3'
5'-GACAGACAGACAGACA-3'
5'-TCCTAACCCTAAATCCAGCTCATGCC-3'
5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-CCCAACCAAGACCTACTTCC-3'
5'-ACTGACTGACTGACTG-3'
5'-TGTCTGTCTGTCTGTC-3'
5'-GCTCCGGAGGAAGGGC-3'
5'-GACACGACACGACACGACAC-3'
5'-CTCCTTCTCCAGCTGC-3'
5'-GAGGGTGGCGGTTCT-3'
5'-GAGGGTGGTGGCTCT-3'
5'-AGAACCGCCACCCTC-3'
5'-GGAGCTGGAGAAGGAG-3'
5'-TGGATGGATGGATGGATGGA-3'
5'-CGAGGCTGCCAACCACATTGATGAC-3'
5'-CACCACCACCACCAC-3'
5'-TTGCCTGTCTCCAGC-3'.

34. A mixture of nucleic acid sequences, comprising a plurality of double-stranded products comprising a first and a second single-stranded polynucleotides each having a 5'-terminal region sequence having at least eleven nucleotides and a 3'-terminal region sequence having at least eleven nucleotides which is substantially the inverted compliment thereto, and an internal region.

35. The mixture of claim 34 wherein at least a portion of each said 5'-terminal region of said polynucleotides are substantially identical for all of said products.

36. The mixture of claim 35 wherein said products can be separated into a finite number of groups according to size.

37. The mixture of claim 36 wherein each of said products of different size contain different sequences in their internal regions.

38. A method for amplifying a plurality of sequences found in a nucleic acid sample comprising:

a) providing a nucleic acid sample comprising nucleic acid sequences from a distinct nucleic acid source;

b) bringing said sample together with a single oligonucleotide primer comprising a sequence having at least eleven nucleotides and capable of hybridizing to a plurality of regions within said nucleic acid sequences, under conditions which synthesize an extension product which is complementary to a portion of said nucleic acid sequences, and wherein said extension product, when separated from said

nucleic acid sequence, is also capable of hybridizing to said oligonucleotide primer;

c) separating the primer extension products from the nucleic acid sequences to which they were hybridized to form single stranded templates; and

d) bringing together the templates of step (c) with the oligonucleotide primer of step (b) under conditions which synthesize a primer extension product from the primers hybridized to said templates, thereby amplifying a plurality of portions of said nucleic acid sequences, so that discrete portions of said nucleic acid sequences are detectable.

39. The method of claim 38 wherein said primer has a sequence of deoxyribonucleotides of the formula:

$$5'\text{-}X_1\text{-}X_2\text{-}X_3\text{-}X_4\text{-}G\text{-}A\text{-}C\text{-}Y_1\text{-}Y_2\text{-}Y_3\text{-}3'$$

wherein

X_1 is $desX_1$ or an oligonucleotide of from 3 to 11 bases selected from the group consisting of A-G-A-G, G-A-C-C-A-A-C-T-G-G-T, and C-C-C;

X_2 is $desX_2$ or an oligonucleotide of 3 bases selected from the group consisting of A-A-T and C-C-C;

X_3 is $desX_3$ or an oligonucleotide of from 1 to 3 bases selected from the group consisting of T-C-T, G-G-T, A-A-C, and G;

X_4 is an oligonucleotide of 3 bases selected from the group consisting of C-A-A, A-G-C and A-C-A;

Y_1 is a base selected from the group consisting of C and A;

Y_2 is an oligonucleotide of 3 bases selected from the group consisting of G-G-C, C-C-T, G-A-C, and T-A-C; and

Y_3 is desX_3 or an oligonucleotide of from 5 to 13 bases selected from the group consisting of A-A-C-A-G-G, G-G-G-C-C-T-G-G-T-C-G-A-T, A-G-A-C-A and T-T-C-C-C-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

40. The method of claim 39 wherein said sequence has the formula:

5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-Y_2-Y_3-3'

wherein X_1, X_2, X_3, Y_2 and Y_3 are as previously defined.

41. The method of claim 39 wherein said sequence has the formula:

5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-G-(G/A)-C-Y_3-3'

wherein X_1, X_2, X_3 and Y_3 are as previously defined.

42. The method of claim 39 wherein said sequence is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'

5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'

5'-GACCAACTGGTAATGGTAGCGACCGGC-3'

5'-GACAGACAGACAGACA-3' and

5'-CCCCCCAACCAAGACCTACTTCCCCC-3'.

43. The method of claim 38 wherein said primer has a sequence of deoxyribonucleotides of the formula:

5'-W_1-W_2-W_3-T-G-G-(T/G)-G-G-(C/G)-Z_1-3'

wherein

W_1 is an oligonucleotide of from 7 to 12 bases selected from the group consisting of G-G-G-G-G-A-A-G-T-A-G-G, G-T-C-C-A-T-C-A-A, and A-G-C-G-A-G-G;

W_2 is an oligonucleotide of 2 bases selected from the group consisting of A-T and T-C;

W_3 is a base selected from the group consisting of T, G, and A; and

Z_1 is an oligonucleotide of from 3 to 7 bases selected from the group consisting of G-G-G, A-G-C-C-T-C-G and T-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

44. The method of claim 43 wherein said sequence has the formula:

$5'-W_1-A-T-W_3-T-G-G-T-(T/G)-G-G-(C/G)-Z_1-3'$

wherein W_1, W_3, and Z_1 are as previously defined.

45. The method of claim 43 wherein said sequence has the formula:

$5'-W_1-A-T-W_3-T-G-G-T-T-G-G-C-Z_1-3'$

wherein W_1, W_3, and Z_1 are as previously defined.

46. The method of claim 43 wherein said sequence is selected from the group consisting of:

5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'

5'-GTCCATCAATGTGGTTGGCAGCCTCG-3' and

5'-AGCGAGGATATGGTGGGCTCC-3'.

47. The method of claim 38 wherein said primer is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'
5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'
5'-GACCAACTGGTAATGGTAGCGACCGGC-3'
5'-ATGGCCTTCCAAAACGACGTCTA-3'

5'-CAAGACCGGCAACAGGATTC-3'
5'-TGGAGGAAGGGCTGGAGGAGGGCTCCGGAGGAAGGGC-3'
5'-GCCCTTCCTCCGGAGCCCTCCTCCAGCCCTTCCTCCA-3'
5'-GAGGTGGGCAGGTGGA-3'
5'-GACAGACAGACAGACA-3'
5'-TCCTAACCCTAAATCCAGCTCATGCC-3'
5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-CCCAACCAAGACCTACTTCC-3'
5'-ACTGACTGACTGACTG-3'
5'-TGTCTGTCTGTCTGTC-3'
5'-GCTCCGGAGGAAGGGC-3'
5'-GACACGACACGACACGACAC-3'
5'-CTCCTTCTCCAGCTGC-3'
5'-GAGGGTGGCGGTTCT-3'
5'-GAGGGTGGTGGCTCT-3'
5'-AGAACCGCCACCCTC-3'
5'-GGAGCTGGAGAAGGAG-3'
5'-TGGATGGATGGATGGATGGA-3'
5'-CGAGGCTGCCAACCACATTGATGAC-3'
5'-CACCACCACCACCAC-3'
5'-TTGCCTGTCTCCAGC-3'.

48. A primer comprising an oligonucleotide sequence having at least eleven nucleotides and capable of hybridizing to a plurality of regions within nucleic acid sequences contained in a sample of nucleic acids, said oligonucleotide primer being capable of hybridizing to an extension product which contains said oligonucleotide sequence and is substantially complementary to a portion of said nucleic acid sequences.

49. The primer of claim 48 wherein said primer has a sequence of deoxyribonucleotides of the formula:

$$5'-X_1-X_2-X_3-X_4-G-A-C-Y_1-Y_2-Y_3-3'$$

wherein

X_1 is desX_1 or an oligonucleotide of from 3 to 11 bases selected from the group consisting of A-G-A-G, G-A-C-C-A-A-C-T-G-G-T, and C-C-C;

X_2 is desX_2 or an oligonucleotide of 3 bases selected from the group consisting of A-A-T and C-C-C;

X_3 is desX_3 or an oligonucleotide of from 1 to 3 bases selected from the group consisting of T-C-T, G-G-T, A-A-C, and G;

X_4 is an oligonucleotide of 3 bases selected from the group consisting of C-A-A, A-G-C and A-C-A;

Y_1 is a base selected from the group consisting of C and A;

Y_2 is an oligonucleotide of 3 bases selected from the group consisting of G-G-C, C-C-T, G-A-C, and T-A-C; and

Y_3 is desX_3 or an oligonucleotide of from 5 to 13 bases selected from the group consisting of A-A-C-A-G-G, G-G-G-C-C-T-G-G-T-C-G-A-T, A-G-A-C-A and T-T-C-C-C-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

50. The primer cf claim 49 wherein said sequence has the formula:

5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-Y_2-Y_3-3'

wherein X_1, X_2, X_3, Y_2 and Y_3 are as previously defined.

51. The primer of claim 49 wherein said sequence has the formula:

5'-X_1-X_2-X_3-C-A-A-G-A-C-(C/A)-G-(G/A)-C-Y_3-3'

wherein X_1, X_2, X_3 and Y_3 are as previously defined.

52. The primer of claim 49 wherein said sequence is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'

5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'

5'-GACCAACTGGTAATGGTAGCGACCGGC-3'

5'-GACAGACAGACAGACA-3' and

5'-CCCCCCAACCAAGACCTACTTCCCCC-3'.

53. The primer of claim 48 wherein said primer has a sequence of deoxyribonucleotides of the formula:

$5'-W_1-W_2-W_3-T-G-G-(T/G)-G-G-(C/G)-Z_1-3'$

wherein

W_1 is an oligonucleotide of from 7 to 12 bases selected from the group consisting of G-G-G-G-A-A-G-T-A-G-G, G-T-C-C-A-T-C-A-A, and A-G-C-G-A-G-G;

W_2 is an oligonucleotide of 2 bases selected from the group consisting of A-T and T-C;

W_3 is a base selected from the group consisting of T, G, and A; and

Z_1 is an oligonucleotide of from 3 to 7 bases selected from the group consisting of G-G-G, A-G-C-C-T-C-G and T-C-C;

with the proviso that at least approximately fifty percent (50%) of said deoxyribonucleotides are deoxyguanylic acid and deoxycytidylic acid.

54. The primer of claim 53 wherein said sequence has the formula:

$5'-W_1-A-T-W_3-T-G-G-T-(T/G)-G-G-(C/G)-Z_1-3'$

wherein W_1, W_3, and Z_1 are as previously defined.

55. The primer of claim 53 wherein said sequence has the formula:

$5'-W_1-A-T-W_3-T-G-G-T-T-G-G-C-Z_1-3'$

wherein W_1, W_3, and Z_1 are as previously defined.

56. The primer of claim 53 wherein said sequence is selected from the group consisting of:

5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'

5'-GTCCATCAATGTGGTTGGCAGCCTCG-3' and

5'-AGCGAGGATATGGTGGGCTCC-3'.

57. The primer of claim 48 wherein said primer is selected from the group consisting of:

5'-CAAGACCGGCAACAGG-3'
5'-AGAGAATTCTCAAGACCCCTGCGCCTGGTCGAT-3'
5'-GACCAACTGGTAATGGTAGCGACCGGC-3'
5'-ATGGCCTTCCAAAACGACGTCTA-3'
5'-CAAGACCGGCAACAGGATTC-3'
5'-TGGAGGAAGGGCTGGAGGAGGGCTCCGGAGGAAGGGC-3'
5'-GCCCTTCCTCCGGAGCCCTCCTCCAGCCCTTCCTCCA-3'
5'-GAGGTGGGCAGGTGGA-3'
5'-GACAGACAGACAGACA-3'
5'-TCCTAACCCTAAATCCAGCTCATGCC-3'
5'-GGGGGAAGTAGGTCTTGGTTGGGGGG-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-GTCATCAATGTGGTTGGCAGCCTCG-3'
5'-AGCGAGGATATGGTGGGCTCC-3'
5'-CCCAACCAAGACCTACTTCC-3'
5'-ACTGACTGACTGACTG-3'
5'-TGTCTGTCTGTCTGTC-3'
5'-GCTCCGGAGGAAGGGC-3'
5'-GACACGACACGACACGACAC-3'
5'-CTCCTTCTCCAGCTGC-3'
5'-GAGGGTGGCGGTTCT-3'
5'-GAGGGTGGTGGCTCT-3'
5'-AGAACCGCCACCCTC-3'
5'-GGAGCTGGAGAAGGAG-3'
5'-TGGATGGATGGATGGATGGA-3'
5'-CGAGGCTGCCAACCACATTGATGAC-3'
5'-CACCACCACCACCAC-3'
5'-TTGCCTGTCTCCAGC-3'.

FIGURE 1

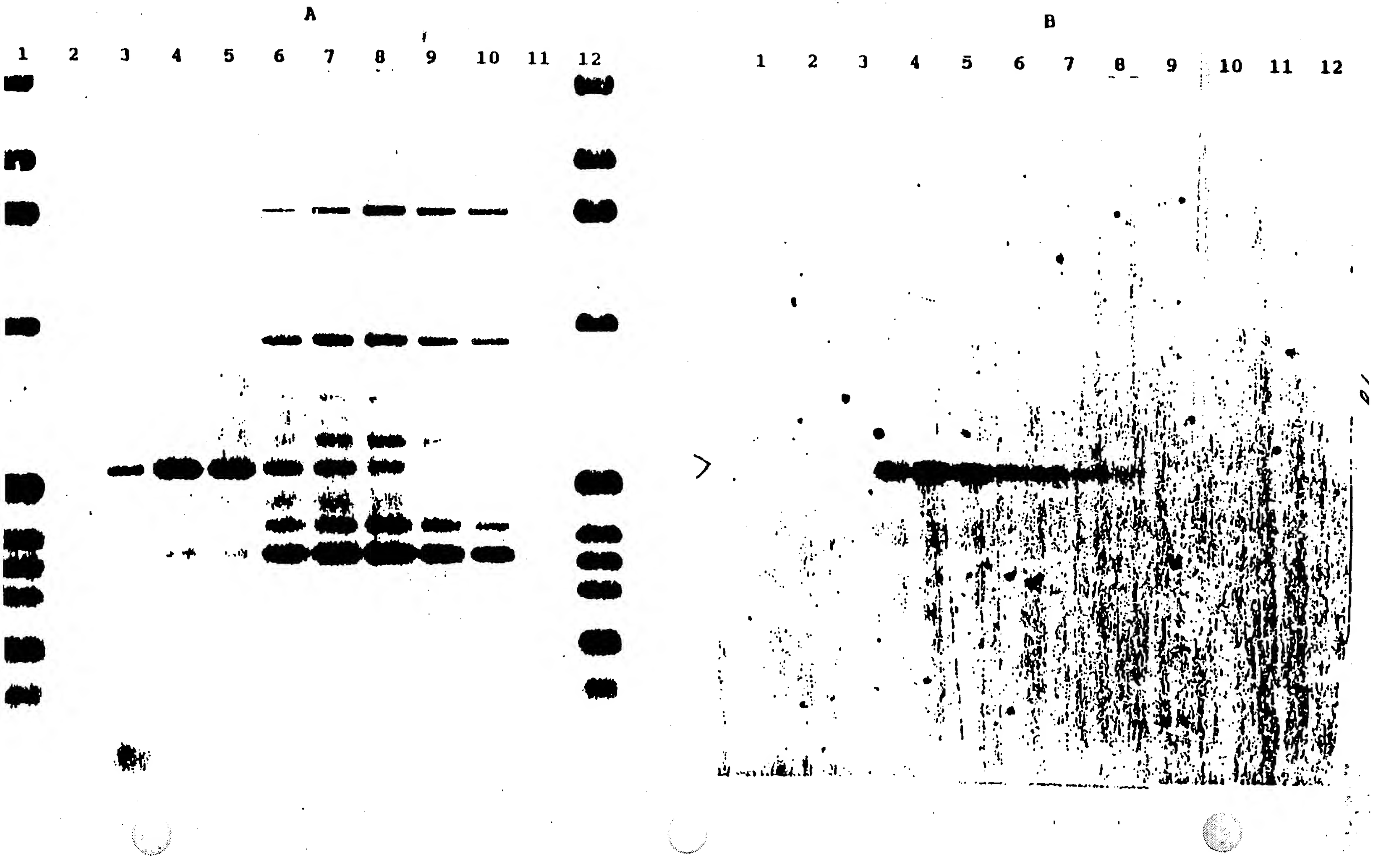
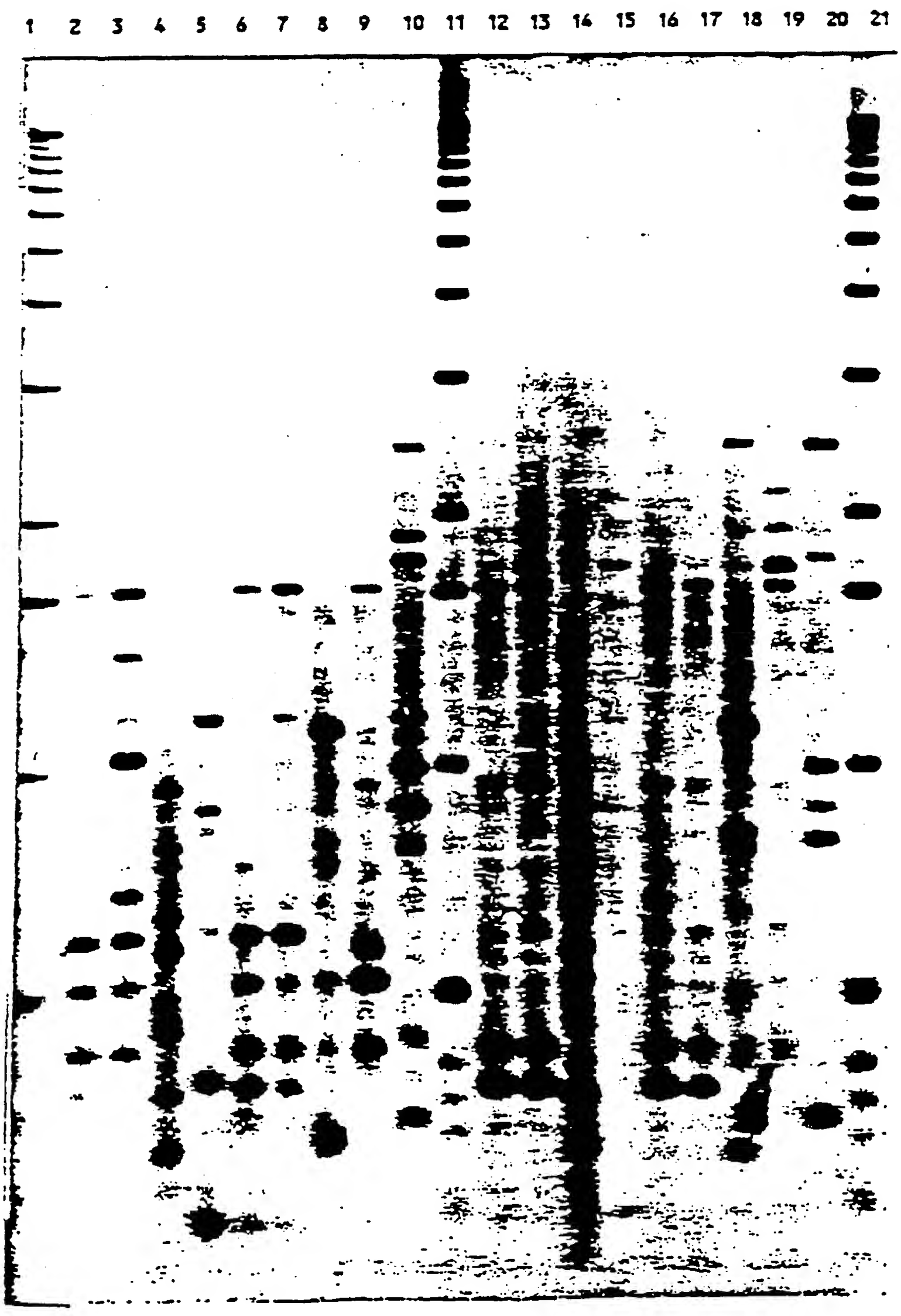


FIGURE 2



3/5

FIGURE 3A

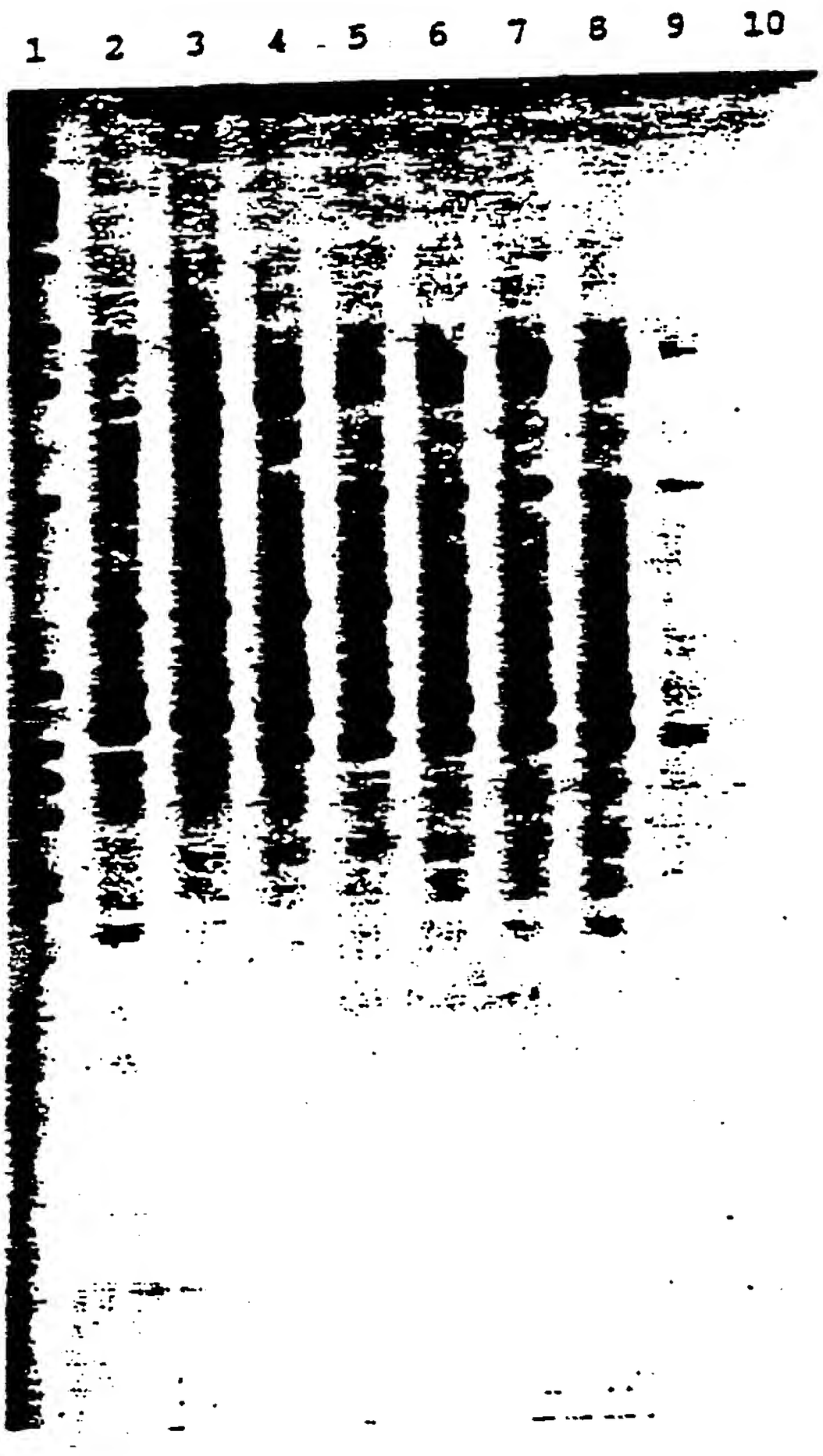


FIGURE 3B

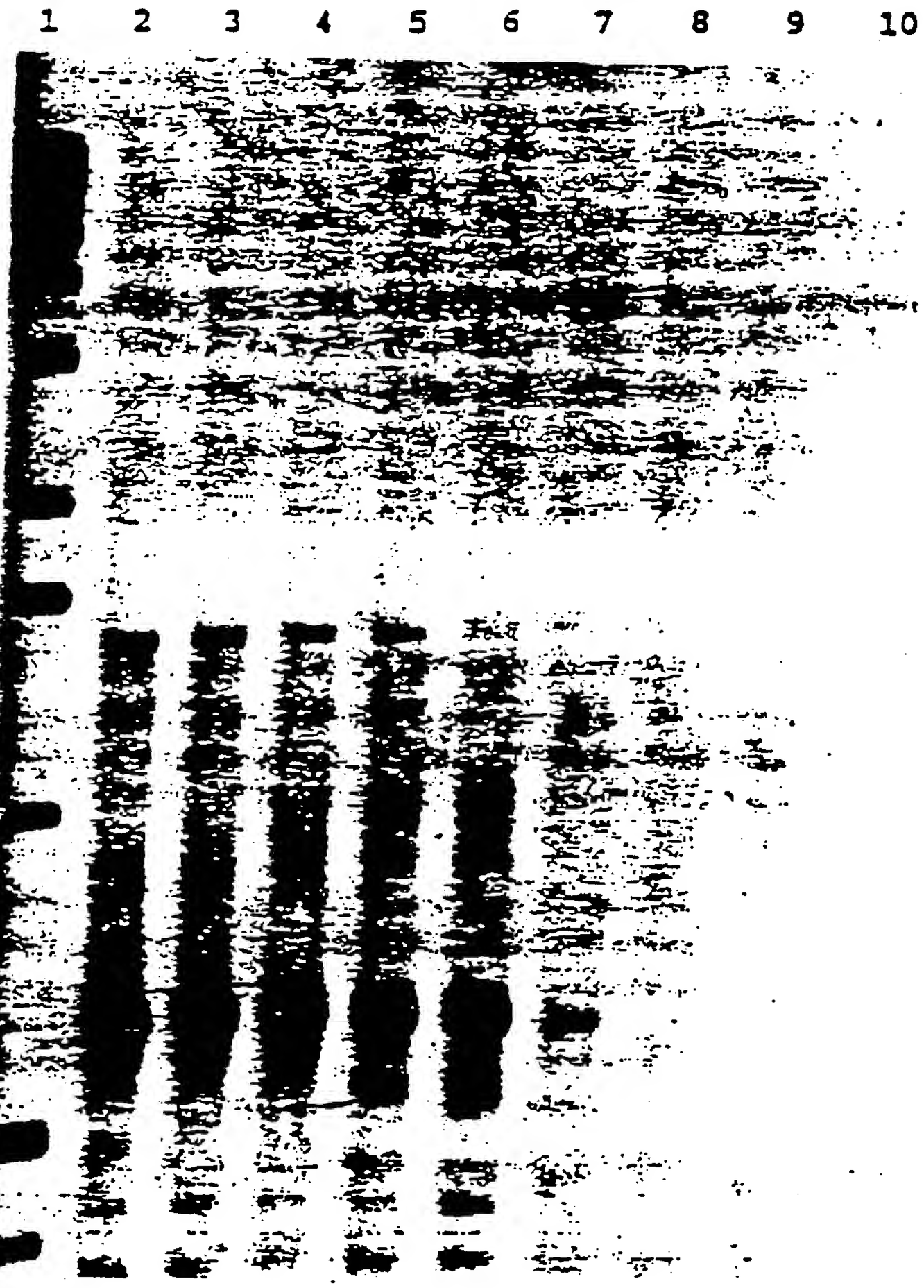


FIGURE 3C

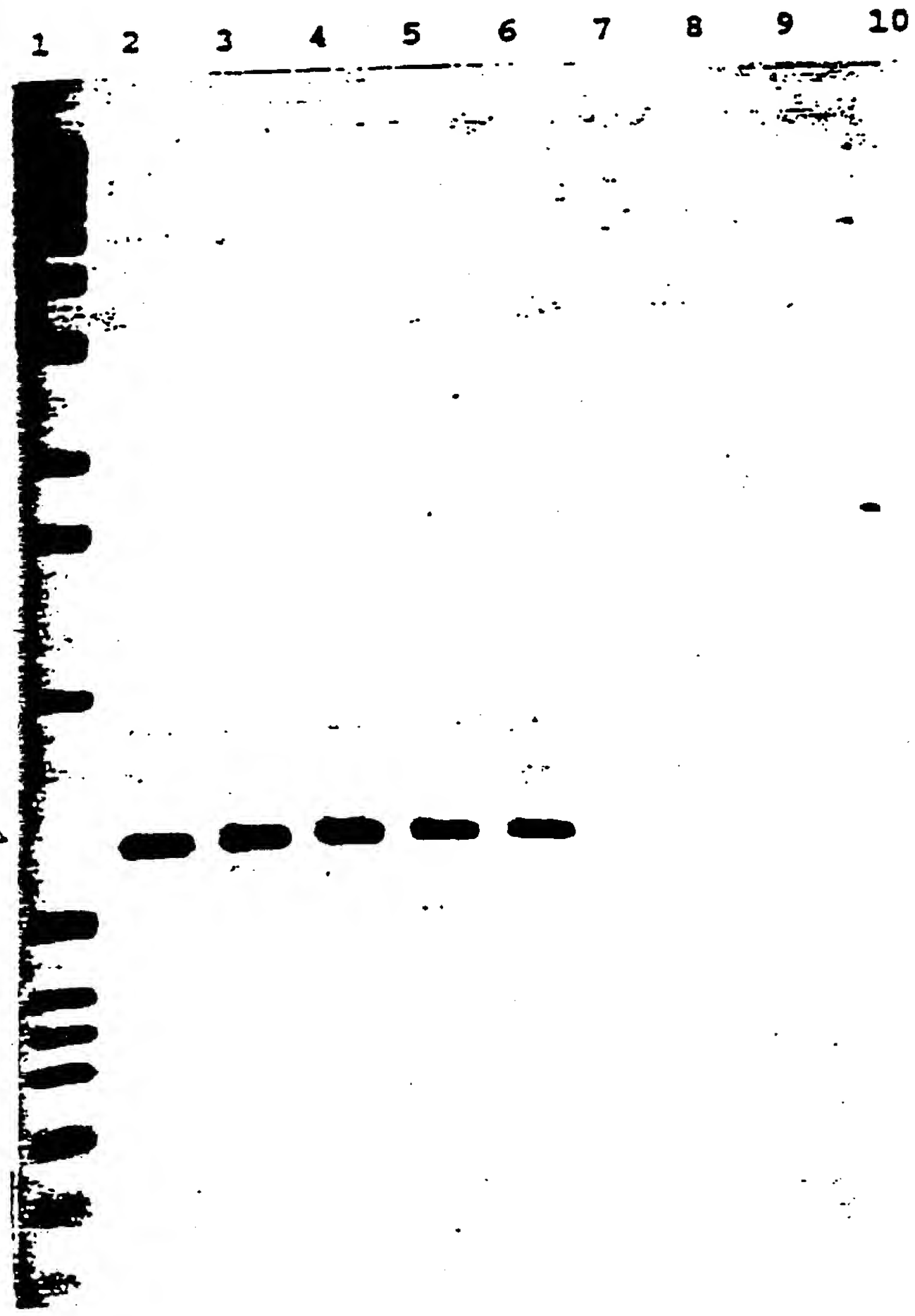
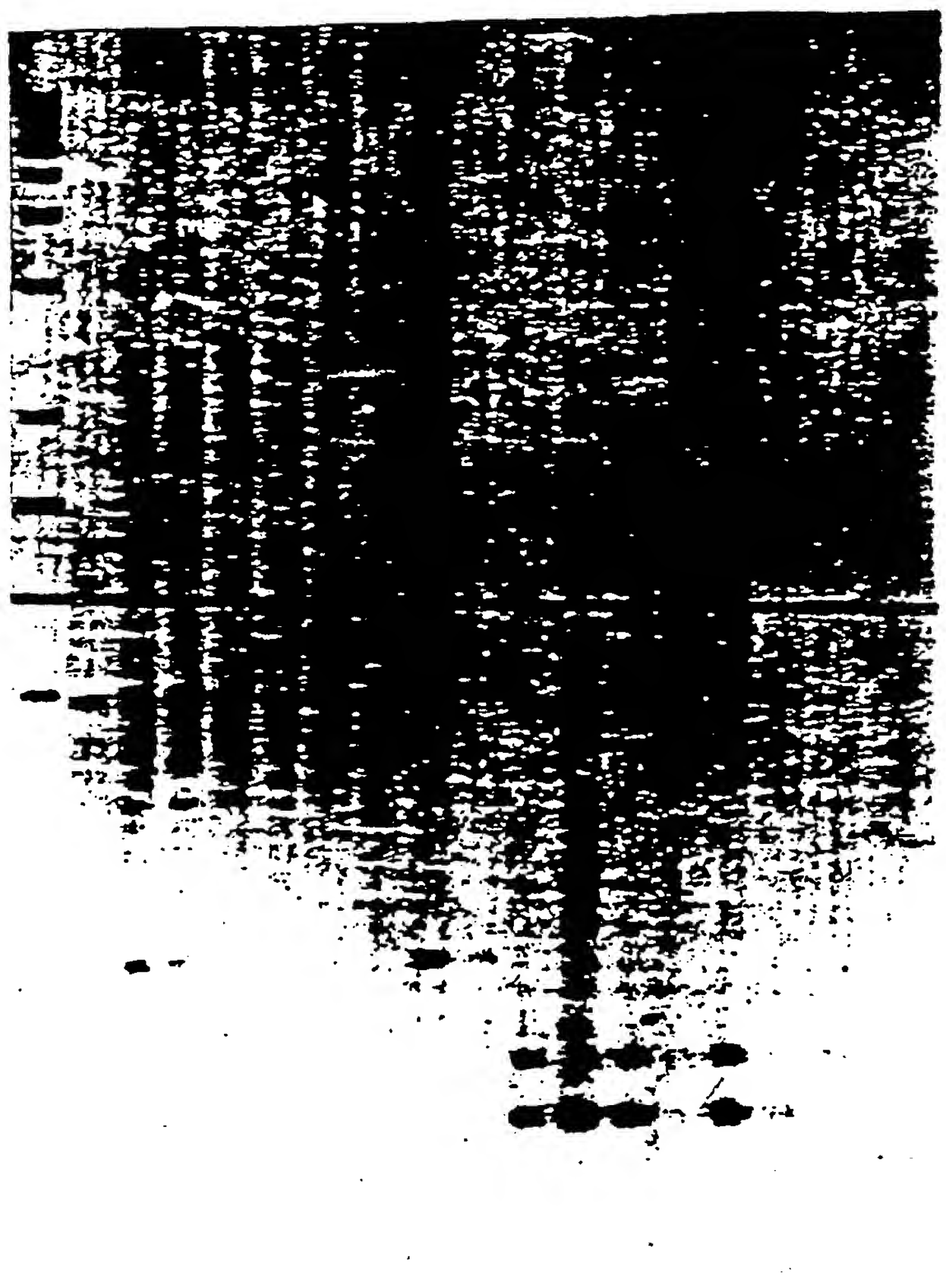
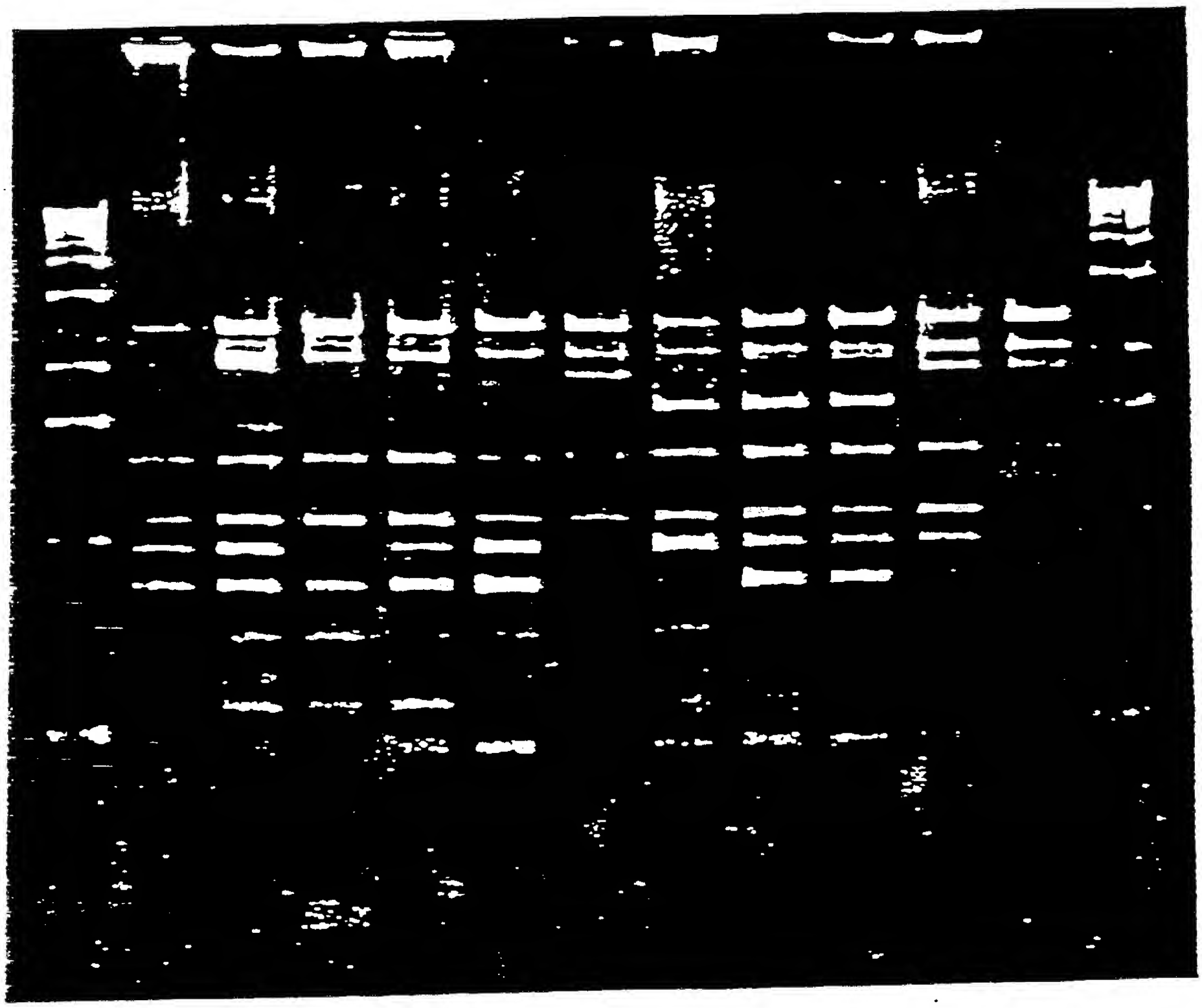


FIGURE 4



8/8

FIGURE 5B



FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

A	WO, A, 86/02948 (JEFFREYS) 22 May 1986. see entire disclosure.	1-57
X/Y	Proceedings of the National Academy of Sciences. Volume 86, issued September 1989, Nelson et al., "Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources", pages 6686-6690, see especially the abstract.	1,2,4,5,17,19 21,22,34-38,48 3,6-16,18,20, 23-33,39-47,49-57

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers . because they relate to subject matter [2] not required to be searched by this Authority, namely:

2. ☐ Claim numbers . because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [3], specifically:

3. ☐ Claim numbers _____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [4]

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

III DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X Y	Nucleic Acids Research, Volume 16, Number 23, issued 1988, Jeffreys et al., "Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells," pages 10953-10971, see especially the abstract.	1,2,4,5,17,19,21,22,34-38,48 3,6-16,18,20,23-33,39-47,49-57
X Y	Proceedings of the National Academy of Sciences, Volume 85, issued December 1988, Frohman et al., "Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer", pages 8998-9002, see especially the abstract and Figure 1 on page 8999.	1-7,17,19-24,34-38,48 6-16,18,25-33,39-47,49-57
X Y	Gene, Volume 84, issued February 1989, Shyamala et al., "Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR", pages 1-8, see especially the abstract and Figure 1 on page 3.	1-7,17,19-24,34-38,48 6-16,18,25-33,39-47,49-57
X Y	Biochemical and Biophysical Research Communications, Volume 167, Number 2, issued 16 March 1990, Kalman et al., "Polymerase chain reaction (PCR) amplification with a single specific primer", pages 504-506, see especially the abstract and Figure 1 on page 505.	1-7,17,19-24,34-38,48 6-16,18,25-33,39-47,49-57

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X/Y	Genomics, Volume 6, issued 1990, Ledbetter et al., "Rapid Isolation of DNA Probes Within Specific Chromosome Regions by Interspersed Repetitive Sequence Polymerase Chain Reaction", pages 475-481, see especially the abstract	1,2,4,5,17,19 21,22,34-38,48 / 3,6-16,18,20,23-33,39-47,49-57
A	GB, A, 2,166,445 (JEFFREYS) 08 May 1986, see entire disclosure.	1-37
X/Y	Journal of Experimental Medicine, Volume 167, issued January 1988, Shibata et al., "Detection of Human Papilloma Virus in Paraffin-Embedded Tissue Using the Polymerase Chain Reaction", pages 225-230, see especially the Summary section bridging pages 229 and 230.	1-3,17-20,34-38 48 / 4-16,21-33,39-47, 49-57
X/Y	Science, Volume 230, issued 20 December 1985, Saiki et al., "Enzymatic Amplification of B-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia", pages 1350-1354, see especially the abstract.	1,2,4,5,17,19, 21,22,34-38,48 / 3,6-16,18,20, 23-33,39-47,49-57
A	NATURE, Volume 314, issued 07 March 1985, Jeffreys et al., "Hypervariable minisatellite regions in human DNA, "pages 67-73, see entire disclosure.	1-57
A	WO, A, 84/04758 (HELENTIARIS ET AL.) 06 December 1984, see entire disclosure.	1-57